



Published in final edited form as:

*Arch Gen Psychiatry*. 2012 September ; 69(9): 913–924. doi:10.1001/archgenpsychiatry.2012.65.

## Towards clinically useful neuroimaging in depression treatment: Is subgenual cingulate activity robustly prognostic for depression outcome in Cognitive Therapy across studies, scanners, and patient characteristics?

Greg J. Siegle, Ph.D.<sup>1</sup>, Wesley K. Thompson, Ph.D.<sup>2</sup>, Amanda Collier, B.S.<sup>1</sup>, Susan R. Berman, M.Ed.<sup>1</sup>, Joshua Feldmiller, B.A.<sup>1</sup>, Michael E. Thase, M.D.<sup>3</sup>, and Edward S. Friedman, M.D.<sup>1</sup>

<sup>1</sup>University of Pittsburgh School of Medicine

<sup>2</sup>University of California, San Diego

<sup>3</sup>University of Pennsylvania School of Medicine

### Abstract

**Context**—40–60% of unmedicated depressed individuals respond to Cognitive Therapy (CT) in controlled trials. Multiple previous studies suggest that activity in the subgenual anterior cingulate predicts outcome in CT for depression, but there have been no prospective replications.

**Objective**—This study prospectively examined whether subgenual cingulate activity is a reliable and robust prognostic outcome marker for CT for depression and whether its activity changes in treatment.

**Design**—Two inception cohorts were assessed with fMRI on different scanners on a task sensitive to sustained emotional information processing before and after 16–20 sessions of CT, along with a sample of control participants tested at comparable intervals.

**Setting**—Therapy took place in a hospital outpatient clinic.

---

Corresponding author: Greg Siegle, Ph.D., Western Psychiatric Institute and Clinic, 3811 O'Hara St., Pittsburgh, PA 15213, ph: 412-383-5444, fax: 412-383-5426, gsiegle@pitt.edu.

**Previous Presentation.** Portions of this manuscript were presented at the meeting of the World Congress for Behavioral and Cognitive Therapies (2010), Boston, MA.

**Location of Work:** University of Pittsburgh, School of Medicine

#### Ethics

Associated studies were approved by the University of Pittsburgh Institutional Review Board (IRB). After being fully informed regarding the nature of the study, all participants gave written consent to participate in the study by signing University of Pittsburgh.

**Disclosures:** No authors had conflicts relevant to this manuscript. Greg Siegle is an unpaid consultant for Trial IQ and Neural Impact. During the past three years, Michael E. Thase reported having served as an advisor/consultant for Alkermes, AstraZeneca, Bristol-Myers Squibb Company, Eli Lilly & Co., GlaxoSmithKline, Janssen Pharmaceutica, MedAvante, Inc., Merck Inc., Neuronetics, Inc., Novartis, Inc., Otsuka Inc., PamLab Inc., Pfizer Inc., Pharmaneuroboost Inc., Shire US Inc., Supernus Pharmaceuticals, and Wyeth Pharmaceuticals. In the two years prior to 2011, he had Speakers Bureau affiliations with AstraZeneca, Bristol-Myers Squibb Company, Eli Lilly & Co., Merck Inc., and Wyeth Pharmaceuticals. He has equity holdings in MedAvante, Inc., has received royalties from American Psychiatric Publishing, Inc., Guilford Publications, and Herald House, and has a family relationship with senior staff at Embryon Inc (Formerly Cardinal Health and Advogent). During the past three years, Dr. Thase has received research funding from Agency for Healthcare Research and Quality, Eli Lilly and Co., GlaxoSmithKline, the National Institute of Mental Health, Otsuka Inc., Pfizer Inc., Pharmaneuroboost, and Roche Inc. Dr. Friedman has provided Speaker Bureaus or Advisory Boards: AstraZeneca, Eli Lilly GlaxoSmithKline, Pfizer Wyeth-Ayerst, and has obtained Grant/Research support from Aspect Medical Systems, Indevus, AstraZeneca, Bristol-Myers Squibb, Pfizer, Sanofi-Aventis, Wyeth-Ayerst, Cyberonics, Novartis, NorthStar/St. Jude Medical, Medtronic, Respiroics

**Patients**—Participants included 49 unmedicated depressed adults and 35 healthy control participants.

**Main Outcome Measures**—Pre-treatment subgenual anterior cingulate activity in an *a priori* region in response to negative words was correlated with residual severity and used to classify response and remission.

**Results**—As expected, in both samples, participants with the lowest pre-treatment sustained subgenual cingulate (sgACC; BA25) reactivity in response to negative words displayed the most improvement in CT ( $R^2=.29$ , >75% correct classification of response, >70% correct classification of remission). Other *a priori* regions explained additional variance. Response/Remission in Cohort 2 was predicted based on thresholds from Cohort 1. sgACC activity remained low for remitters following treatment.

**Conclusions**—Neuroimaging provides a quick, valid, and clinically applicable way of assessing neural systems associated with treatment response/remission. sgACC activity, in particular, may reflect processes which interfere with treatment, e.g., emotion generation in addition to its putative regulatory role; alternately, its absence may facilitate treatment response.

## Keywords

Cognitive Therapy; Mood Disorders – Unipolar; Brain Imaging Techniques; Cognitive Neuroscience; Emotion

## Introduction

Cognitive Therapy (CT) <sup>1</sup> is a common empirically supported intervention which addresses systematic negative thinking and is effective for 40–60% of patients with unipolar depression <sup>2</sup>. Knowing which patients are likely to benefit from CT could increase response rates and decrease costs by targeted referrals. Yet, biological predictors of CT outcome e.g., <sup>3, 4–6</sup> have not been adopted clinically. This may be due to the lack of reliability and validity data for these measures or lack of ability to overcome variability across scanners/labs. Here we examined whether our previously-observed association of CT outcome with pre-treatment activity in the subgenual anterior cingulate cortex (sgACC) <sup>4</sup> replicates in multiple new samples and scanners.

The sgACC is an intuitive target to examine as a treatment predictor as it has connections to limbic regions such as the amygdala and has been suggested to serve as a proximal regulator of limbic function <sup>7</sup>. It further has abnormalities in activity in depressed individuals e.g., <sup>8, 9</sup>, changes in CT and treatment with medications <sup>10, 11</sup>, as well as associations with symptom change in depression e.g., <sup>12, 13</sup>. Additionally, it is cytoarchitecturally uniform and has easily anatomically identifiable boundaries making results likely to generalize across studies. It is prognostic for clinical change in multiple CT studies using different paradigms in depression <sup>4, 5</sup> and other disorders such as PTSD <sup>14</sup>.

To pave the way for clinical adoption of this prognostic biomarker, we examined whether sgACC prediction of outcome could be replicated in an efficacy sample (i.e., best possible conditions for observing the effect) as well as an effectiveness sample (i.e., more real-world, less ideal conditions). Thus we examined whether sgACC prediction of depression outcome in CT 1) holds in an efficacy cohort using effectively the same recruitment, design (including task), therapists, sample size, trial selection (e.g., measuring reactivity to negative words), data preparation, and analytic techniques as in <sup>4</sup>; in addition to common arguments regarding the importance of replication before translation to the clinic, particularly given a single N=14 study, the need for strict replication is particularly great for voxelwise neuroimaging studies as relationships of voxelwise activity to self-report measures are

notoriously unreliable<sup>15, 16</sup>, 2) holds in an effectiveness cohort with more heterogeneous clinically representative sample and community therapists with variable supervision, 3) is revealed as one of the measured neural indices which best predict treatment outcome, 4) can be formulated in a way that can be easily interpreted by clinicians and used across scanners, and 5) changes in treatment for those with the predictive marker, to inform, if not confirm, causal inferences regarding the possible role of sgACC change in clinical change.

For prediction (goals 1–2), our primary hypothesis was that sgACC activity would be prognostic for outcome in CT in multiple samples, robustly enough to garner clinical consideration. Our secondary hypotheses were that other relevant theoretically motivated regions and their functional relationships would add variance to prediction of outcome but would not completely moderate the role of the sgACC (goal 3), that the same predictive associations would be apparent in normalized data (goal 4), and that sgACC activity would normalize following treatment (goal 5).

For the “other relevant regions” (goal 3), we examined a brain network associated with emotional reactivity and implicated in depression and treatment outcome<sup>17</sup> including the left amygdala as in<sup>4, 18</sup> and regions associated with regulatory control that have decreased functioning in depression: the dorsolateral prefrontal cortex e.g.,<sup>8, 19, 20</sup>, rostral anterior cingulate<sup>21–23</sup>, and their functional corticolimbic/corticocortical relationships e.g.,<sup>24, 25</sup>, as well as a voxelwise analysis.

Directional hypotheses for change in sgACC activity (goal 5) are dependent on the somewhat ambiguous function of this region. On the theory that the sgACC inhibits limbic regions e.g.,<sup>26, 27, 28</sup>, we have suggested that as CT teaches skills for emotion regulation, individuals who most need CT, i.e., those with the lowest pretreatment levels of sgACC activity, may respond best to CT<sup>4</sup>. Thus, CT remitters would be expected to show decreased pre-treatment but increased post-treatment sgACC activity. Alternately, the sgACC may also support emotion generation / monitoring. In support of this theory, 1) neurofeedback-induced increased sgACC activity yields increased sadness<sup>29</sup>, 2) trait-related increased sgACC is associated with higher sadness<sup>30</sup> and increased depressive severity<sup>31</sup>, and 3) sgACC inhibition via deep brain stimulation facilitates recovery e.g.,<sup>32</sup>. If this role is primary, low sgACC activity may be needed for voluntary regulation to occur. In this case, remitters would be expected to show decreased pre- and post-treatment sgACC activity.

To evaluate whether sgACC activity “normalizes” in treatment, a sample of healthy never-depressed control participants were recruited to establish a “normative” baseline. The controls also afforded the ability to examine pretreatment activity with respect to normative function, i.e., as clinically interpretable Z-scores (goal 4) and to assess the test-retest reliability outside the context of depression – i.e., showing that healthy individuals do not change strongly over time and whether the measure is reliable enough to make inferences regarding pre-post measurements (e.g., as recommended in<sup>33</sup>). Data from controls was not used in primary prediction analyses.

## Method

### Participants and therapists

As shown in Figure 1 (CONSORT diagram), and Table 1 (demographics), participants from two clinical trials were tested in two cohorts on different scanners, separated by approximately 6 months (full elaboration in Author Material-I). After attrition and data cleaning (fully described in Figure 1), Cohort 1 (efficacy: from Thase, PI; ClinicalTrials.gov: NCT00183664) included 17 patients with recurrent major depressive disorder (diagnoses via SCID<sup>34</sup>) treated by the same three therapists from<sup>4</sup> who continued

to receive weekly supervision with tape-review from the same master clinician (Sandar Kornblith, Ph.D.), as well as 15 healthy controls (no current or historical Axis I disorder via SCID interview). Therapeutic adherence was deemed adequate with a random selection of 10 tapes (ratings of  $\geq 2$  tapes per therapist by other therapists or outside raters) receiving high marks on the Cognitive Therapy Rating Scale<sup>35</sup> (scores of 40 represent high levels of adherence), Mean (Stdev)=52.7(8.6), Range = 40–63.

Cohort 2 (effectiveness) included 32 patients who were more clinically heterogeneous than Cohort 1 as they included N=23 with recurrent major depressive disorder but also N=9 in their first-episode. Participants were drawn from the same trial or Siegle, PI, ClinicalTrials.gov: NCT00787501 (CT election given options of CT or SSRI treatment). Cohort 2 also included 20 healthy controls (no current or historical Axis I disorder via SCID interview). Patients were treated by 6 community clinicians (Ph.D.'s, M.D.'s, M. Ed.'s, LCSW's) who ranged in CT experience from a Ph.D. founding member of the Academy of Cognitive Therapy to a social worker who took her second CT case as part of this study. Therapists received group supervision monthly without tape review, and as-requested supervision by Dr. Kornblith (weekly for 2 therapists). Therapeutic adherence was more variable with 12 tapes of 4 therapists receiving Cognitive Therapy Rating Scale scores of Mean (Stdev)=43.3(13.1), Range=23–61 with 6/12 tapes falling below the cutoff of 40 for adequate adherence, as administered by other study therapists or outside raters.

Participants described no health problems, eye problems, or psychoactive drug abuse in the past six months and no history of psychosis, manic, or hypomanic episodes. Neither control nor depressed participants had used antidepressants within two weeks of testing (six weeks for fluoxetine) due to either medication naivety or supervised withdrawal from unsuccessful medications. Participants reported no excessive use of alcohol in the two weeks prior to testing and scored in the normal range on a cognitive screen,<sup>36</sup> VIQ-equivalent > 85.

## Procedure

After complete study description, IRB-approved written informed consent was obtained followed by a SCID interview, vision test, and unrelated physiological assessment. Patients were assessed on a different day with a battery of fMRI tasks administered in counterbalanced order; one task is reported here (others described in Author Material III-A). Participants rated their sad, anxious, and happy affect from 1 (not at all) to 5 (very) before and after the task. The Beck Depression Inventory II<sup>37</sup> (BDI) was administered following fMRI to assess depressive severity (rationale in Author Material II-A). Depressed participants then received CT; 2 sessions/week for the first four weeks followed by 8 weekly sessions for “early-responders” (HRSD reduction <40% at session 9; 16 total sessions) or 2 sessions/week for the first 8 weeks followed by 4 weekly sessions for non-early-responders (20 total sessions). CT followed Beck's<sup>1</sup> guidelines (detail in Author Material-I) with weekly Hamilton Rating Scale for Depression (HRSD) ratings. Within two weeks of completion (week 16 for controls) all participants completed the same fMRI protocol and BDI again.

## Clinical data preparation

In Cohort 1, 1 participant's pre-treatment and 1 participant's post-treatment BDI were missing. Fortunately the BDI-I was administered clinically in the protocol, and thus these values were reconstructed via regression (described Author Material-II-B). In all other cases, final BDI scores were assessed at participants' second scan (after completing CT or approximately 12 weeks from their first scan for those who did not complete). Final HRSD scores were imputed based on the trajectory of weekly responses as described in Author Material II-D. Response was defined as a 50% reduction in initial BDI or HRSD score (as

opposed to the “early response” criterion of 40% reduction used to determine the number of sessions), and remission was defined as final-BDI <10 (rationale in Author Material-II-C) or final-HRSD <7 (as in STAR\*D).

### **fMRI task and processing**

**Apparatus**—Twenty-nine 3.2mm slices were acquired parallel to the AC-PC line using a reverse direction EPI pulse sequence to minimize susceptibility artifacts in the amygdala and orbitofrontal regions (3T Siemens Trio, T2\*-weighted images depicting BOLD contrast; TR=1500ms, TE=27ms, FOV=24cm, flip=80), yielding 8 whole-brain images per 12 second trial. Stimuli were displayed in black on a white background via a back-projection screen (.88° visual angle). Responses were recorded using a Psychology Software Tools™ glove.

**Personal relevance rating task (PRRT)**—As in our previous publications<sup>4, 20</sup>, in 60 slow-event related trials, participants viewed a fixation cue (1 s; row of X’s with prongs around the center X) followed by a positive, negative, or neutral word (200 ms; only negative words analyzed here), followed by a mask (row of X’s; 10.8 s). Participants pushed a button for whether the word was relevant, somewhat relevant, or not relevant to them or their lives (button orders balanced across participants), as quickly and accurately as they could. Participant-generated and normed words from<sup>38</sup> were used as in our previous studies of depression<sup>4, 20, 39–41</sup> (procedures in Author Material-III; RT preparation in VI-B).

**fMRI data preparation**—Following standard preprocessing similar to that used in our previous study<sup>4</sup> with slight modernizations (slice time correction, motion correction, linear detrending, voxelwise outlier rescaling, conversion to percent-change, temporal smoothing (5 point middle peaked filter), 32 parameter nonlinear warping the Montreal Neurological Institute Colin-27 brain, and spatial smoothing (6mm FWHM), response time-series variability normalization across scanners; methods fully described and compared to<sup>4</sup> in Author Material-IV), the same “reactivity” index used in<sup>4</sup> yielded peak and sustained responses to negative words as the mean of the 4<sup>th</sup>–7<sup>th</sup> images (henceforth “scans”) of each negative-word trial minus the trial’s first (pre-stimulus) scan acquired while the fixation cue was on the screen.

Primary hypotheses were examined for mean reactivity in a 24 voxel 20mm radius sphere centered at Talairach 6, 17, –6, the centroid of the sgACC region prognostic for outcome in<sup>4</sup>. Our broader network included right and left dorsolateral prefrontal cortex (DLPFC) regions, the left amygdala, and BA24 in the rostral-cingulate (criteria described Author Material-V) and to explicitly replicate methods from<sup>4</sup> we also examined regions from voxelwise analyses, subject to empirical Type I error control (Author Material IV-B).

### **Data Analysis: Associations between pre-treatment brain function and clinical change**

Using residual severity controlled for initial severity. To replicate methods from<sup>4</sup> we examined associations of pre-treatment activity with residual severity, computed by regressing final Beck Depression Inventory scores on initial scores and retaining the residuals in participants who completed treatment in Cohort 1. To increase generalizability we then included participants who did not complete treatment in this cohort as well as Cohort 2. To predict response and remission, a within-sample grid-search found the cutoff that maximized %-correct discrimination. Standard indices of signal detection are reported (sensitivity, specificity,  $d'$ ; ROC statistics in Author Material-VIII). 1000 permutation tests using the same algorithm (randomly permuting associations of response or remission with fMRI indices) assessed the significance of discrimination (%-correct and  $d'$ ; tests of  $d'$  generally agreed with %-correct, and thus are only reported when results diverged on significance). To increase generalizability outside clinical trials in which residual severity



can be calculated, we examined Z-scores of sgACC activity (previously described reactivity index) normalized with respect to the same index computed for controls' initial assessment, i.e.,  $sgACC_Z = (X - M_{controls}) / sd_{controls}$ , and severity change scores (final – initial). Multiple regions were associated with residual severity using multiple regression.

Robustness was examined via random forest regression and classification<sup>42</sup>, using as features activity from all *a priori* structures (criteria in Author Material V) and their partial mutual information via<sup>43</sup> to capture systemic effects involving functional-connectivity. Regression and classification forests were trained on Cohort 1 using random re-sampling (bagged); a subset of the activation and mutual information measures was selected by maximizing out-of-bag estimates of prediction and classification accuracy on Cohort 1. We tested generalizability by computing prediction and classification accuracy of the Cohort 1 trained algorithms to Cohort 2. P-values were obtained by permuting responses and computing the proportion of times the permuted regression and classification results were as good or better than the un-permuted results.

### Data Analysis: Change

We examined associations of pre- and post-treatment activity via correlation, and change in the time-series of responses for participants with each combination of predicted and actual response.

## Results

### Demographic and Behavioral Data

Patients and controls did not differ significantly on demographic variables between cohorts or groups (Table 1). Group x cohort ANOVAs revealed that patients rated negative words as more negative and more personally relevant, and reacted more slowly than controls (Author Material-VI). Cohort 2 rated negative words as more negative  $t(82)=3.27$ ,  $p=0.002$ ,  $d=0.73$  with no other Cohort main effects. There were no group x cohort interactions on valence or meaning ratings. As group x cohort effects on rts (Author Material-VI) were small and the hemodynamic window began after their likely effects, they were not used in fMRI analyses. There were no significant associations of rt, valence ratings, or personal relevance ratings with residual BDI scores or sgACC activity,  $|r| < .15$ ,  $p > .4$ . Pre-treatment depressive severity (BDI) was uncorrelated with pre-treatment sgACC activity ( $r = -.13$ ). Depressed participants were moderately sad and anxious and minimally happy before and slightly less so after the task whereas controls were comparably minimally sad and anxious and moderately happy before and after the task (Author Material-VI-C, Table S1).

### Clinical Outcome

CT was successful at rates at least as high as those observed in the literature with enough outcome variability ( $N > 5$  in all response/remission cells) to allow further analysis (Table 1, individual trajectories, Author Material-VII).

**1) Efficacy replication: sgACC activity as prognostic indicator of clinical change in CT**—As shown in Figure 2 and Table 2A, as in our previous study<sup>4</sup>, decreased sgACC activity was strongly correlated with stronger clinical change (more negative  $BDI_{residual}$ ) in Cohort 1, whether or not completers were included. This prediction generalized to the HRSD, which was collected on a slightly larger subset of participants. ROC curves (Author Material-IX) yielded significant area-under-the-curve (AUC), reflecting good discrimination (Author Material-VIII A). Voxelwise associations with  $BDI_{residual}$  revealed sgACC to be associated with even more of the variance ( $> 80\%$  in some

voxels; Figure 3c), suggesting possible optimization and, as in <sup>4</sup>, relative specificity to the sgACC.

## 2) Effectiveness generalization to a more clinically heterogeneous sample with minimally supervised community therapists—

The same general pattern of results held significantly in Cohort 2 though effect sizes were somewhat weaker (Figures 2–3 and Table 2A), potentially due to clinical and therapist heterogeneity. Importantly, prediction of responder and remitter status was still strong with >75% correct predictions for response across all samples and measures, and >65% correct predictions for remission, preserving strong specificity. When data from Cohorts 1 and 2 were combined, yielding a mix of highly supervised, and not-as-strongly supervised therapists, and a wide range of the clinical phenotype, prediction remained strong. ROC analyses (Author Material VIII, IX) were significant reflecting fair discrimination. Remitters were characterized by downward-going hemodynamic responses whereas non-remitters were characterized by upward-going hemodynamic responses (Figure 2E), with significant differences in a 3-group (control, remitter, non-remitter) ANOVA from 6 to 12 seconds,  $F(2,76)=5.563$ ,  $p=0.006$ .

## 3) Do other brain regions or measures explain additional variance?

**Other regions:** A voxelwise mega-analysis combining data across cohorts, (hierarchical regression in which scanner (cohort) and initial BDI were entered on step 1 to simultaneously covary them, with final BDI on step 2), confirmed a 2-voxel region across both samples (Talairach 5, 16, -9) which was both statistically significant at  $p<.001$  ( $p<.05$  corrected) and clinically significant, explaining >40% of the variation in  $BDI_{residual}$ , and located in the *a priori* mask centered around the region detected in our previous study (Figure 3). No other significant >24 voxel regions explaining  $\geq 40\%$  of the variance were detected. Similarly, voxelwise meta-analytic conjunction analyses revealed voxels in the sgACC at  $p<.01$  in each cohort, but no other >24 voxel regions predictive of outcome in both cohorts. Thus, sgACC associations were confirmed and no new regions were added to the *a priori* list.

Associations were similarly strong across all *a priori* regions (Table 2B, Author Material XI) suggesting that the relationship of the sgACC to outcome is not qualitatively unique. Whereas the sgACC explained 29% of the variation, the other 3 regions explained only 12% additional variation, yielding a non-significant increment,  $F(3,38)=2.65$ ,  $p=.06$ , though the final model with all *a priori* regions as predictors was significant, prediction equation:  $BDI_{residual} = -1.80 + 43.13 * sgACC + 17.2 * Amygdala + 35.7 * DLPFC - 12.0 * BA24$ . Thus, decreased possibly regulatory sgACC and DLPFC activity explains better CT outcome.

**Other measures:** Adding common non-fMRI predictors of response including scanner, rumination, pupillary motility during the task, and demographic variables did not explain additional significant variance, and sgACC activity remained a significant predictor (Author Material XI-D).

**Other valences:** Positive and neutral words did not predict residual symptomatology in the combined cohort. With positive, negative, and neutral words in the same model, only negative words were a significant predictor of residual symptomatology (Author Material XIII). Thus, prediction was specific to negative words.

**Robust Classification:** As shown in Table 2C, classification for response and remission status in Cohort 2 was strong based on a sgACC cutoff solely derived from Cohort 1. That is, by relying on the region from a previous study with a cutoff determined in Cohort 1, we achieved 74% correct prediction of response and 78% correct prediction of remission in

Cohort 2. These data suggest that assessment on one scanner yielded a predictive algorithm applicable on another scanner. Adding indices of activity among multiple regions and their connectivity improved classification and continuous prediction of BDI/HRSD<sub>residual</sub> suggesting that accounting for a broader *a priori* network robustly adds to predictive power; had additional terms not statistically improved classification they would not have been retained in the final model. ROC curve analyses (Author Material VIII ROC AUC, Author Material-XI-C) further show that initial severity yielded poor classification (AUC<.7). Multivariate classifications on the test dataset yield significant estimates reflecting fair (AUC>.7) to good (AUC>.8) discriminability with just decreased severity and decreased sgACC and good discriminability with the full model, including decreased sgACC-DLPFC connectivity, predicting outcome.

#### 4) Clinically applicable formulations: Psychometrics, Z- and symptom-change scores

**Reliability and discriminant validity:** sgACC<sub>Z</sub> and sgACC reactivity had moderate test-retest reliability in control participants tested approximately 16 weeks apart (N=27,  $r=.39$ ,  $p=.04$ ) which was similar at each scanner (Cohort 1,  $r=.44$ , Cohort 2,  $r=.35$ ) and did not change when scanner was covaried ( $r=.39$ ). Controls' sgACC mean did not change between the scans,  $t(26)=1.36$ ,  $p=0.19$ ,  $d=0.26$ . All but 1 had  $Z<0.5$  pre-test, and all but 2 at post-test, suggesting stability within a restricted range, decreasing the estimated reliability. Reliability for the composite predictor associated with the sgACC, amygdala, DLPFC, and BA24 was no higher,  $r=.36$ . sgACC<sub>Z</sub> reactivity was minimally correlated with initial depressive severity among depressed patients (N=49)  $r(\text{HRSD})=-.10$ ,  $p=.50$ ;  $r(\text{BDI})=-.21$ ,  $p=.13$ .

**Symptom Change scores:** Associations of sgACC<sub>Z</sub> with individually interpretable BDI<sub>Post-Pre</sub> were nearly identical to associations of %-change with BDI<sub>residual</sub> (Figure 2D, Author Material-VIII, X). For example, in Cohort 1 completers' sgACC<sub>Z</sub> predicted 54% of the variance in BDI<sub>Post-Pre</sub>,  $F(1,15)=16.32$ ,  $p=.001$ , and  $R^2=.29$  in the combined sample  $F(1,42)=16.87$ ,  $p<.005$ . sgACC<sub>Z</sub> cutoffs in the range of .46-.74 strongly predicted response across measures for both samples (76–81% correct response classification), suggesting that high levels of sgACC activity predicted non-response. Similarly, cutoffs closer to 0 predicted remission, so participants with average or higher levels of sgACC responses compared to controls were unlikely to remit in CT.

**5) Towards causal inference: Did sgACC activity change?—**Forty depressed participants had pre and post-treatment fMRI and BDI scores. Those participants with the lowest pre-treatment sgACC activity (primarily CT remitters) also had the lowest post-treatment activity  $r_{\text{pre,post}}=.39$ ,  $F(1,39)=6.82$ ,  $p=0.013$ , (Figure 4A). Depressed participants with pre-treatment activity below the predicted response threshold who remitted had pre- and post-treatment sgACC activity below controls throughout the trial (Figure 4B top left) and did not increase significantly (Table S4, t-test pre. v. post). In contrast, controls with low pre-treatment activity increased significantly (Table S4, Figure S7b) and depressed non-remitters with low pretreatment activity nearly so (Figure 4B top right; statistics Table S4). Moreover, depressed remitters with low pre-treatment activity had a non-significantly smaller proportion of participants who increased and lower mean level of increase than did either controls or non-remitters with low pre-treatment activity (Author Material-XII-B). Thus, we cannot conclude that sgACC activity increased as a function of treatment. Qualitatively, in contrast, 5/6 participants with low-pre-treatment activity who did not remit (Figure 4A, green) showed increased sgACC activity following treatment. Similarly, 3/5 of the remitters who had high pre-treatment sgACC activity decreased post-treatment (Figure 4A, blue squares, average Figure 4B bottom left) whereas 8 of the 12 non-remitters with



high pre-treatment sgACC activity increased (Figure 4A, purple squares). Thus, the emerging picture is that participants with high post-treatment sgACC activity did not remit.

## Discussion

This study replicates and extends previous results<sup>4, 44</sup> suggesting that decreased sgACC reactivity to negative words is prognostic for response, remission, and clinical change in Cognitive Therapy for depression. It was prognostic in both an efficacy sample and a more clinically diverse sample who received more clinically representative care, though effect sizes were somewhat reduced, possibly due to increased patient and therapist heterogeneity and nonstandard administration of Cognitive Therapy. Using measures of brain function interpretable at a single-subject level (Z-scores) and simple change-scores preserved prognostic utility. Response, remission, and change in severity were robustly predicted based on a few *a priori* regions when using thresholds derived in a different sample on a different scanner above and beyond pre-treatment severity. These data suggest that the proposed assessment could have scientific and practical utility. But participants with low pre-treatment sgACC activity who remitted did not generally demonstrate increased sgACC activity suggesting that successful treatment did not operate by normalizing this mechanism but rather remaining more like healthy individuals from pre- to post- treatment.

Though failure to observe an increase in sgACC activity following treatment could be due to the index's low reliability or regression to the mean, these explanations would not explain the association of initial negative-going responses with the smallest sgACC change. Thus, we suggest that increasing sgACC function is not a mechanism of clinical change. Rather, sgACC activity could interfere with voluntary emotion regulation essential for CT, possibly due to its roles in sadness-upregulation<sup>29</sup> or limbic monitoring, possibly via corticocortical connectivity, or automatic down-regulation of emotion<sup>45</sup> preventing the use of considered reappraisal emphasized in CT. These interpretations support the continued development of interventions that decrease sgACC activity e.g.,<sup>32</sup> and do not suggest that CT or other interventions (e.g., neurofeedback) should work to increase ventromedial function. Alternately, low sgACC activity could facilitate CT response if it is easiest to learn to challenge thoughts when emotion-monitoring/generation processes naturally disengage. That multiple regions associated with limbic reactivity and regulation explain overlapping and independent variance suggests that outcome reflects functioning of a wider network.

More practically, because we used *a priori* regions with a single 12-minute fMRI and seven minute structural acquisition, automated preprocessing, and scores that can be calculated on single subjects, our algorithm is feasible and at significantly under \$1000 at most scanning centers, cost effective for use in clinics in a 30 minute MRI appointment without radiologist interpretation. Cross-scanner generalizability is likely because sgACC activity patterns were qualitatively different for remitters (activity decreased from a pre-stimulus baseline) and non-remitters (activity increased) (Table 2, Figure 2E). Thus, inter-scanner scaling differences did not affect prediction. As sensitivity using just the sgACC was low for practical use (Table 2A) but adequate when other regions were accounted for (100% sensitivity in Table 2C generalization set, BDI prediction), it may be useful to assess a network of regions in prediction, or to consider the sgACC primarily as a method to help patients decide what interventions not to try first. As predictions were stronger in the highly controlled cohort, therapists who adhere strongly to a single treatment model may make the most valid use of predictive neuroimaging.

As increased activity in anatomically proximal regions including the rostral cingulate has been shown to positively predict outcome to antidepressants using PET resting state<sup>12, 46, 47</sup>, fMRI reactivity more analogous to the current design<sup>13, 48–52</sup>, and EEG<sup>53</sup>,

ventral cingulate activity could, upon prospective replication, yield an algorithm for selection into CT versus antidepressant treatment<sup>17</sup>. Assessing sgACC function during multi-modal assessments could also help to validate mechanisms for less interpretable prognostic indicators such as psychophysiological assessments<sup>54</sup>. That said, as fMRI-derived sgACC activity explained variance independent of other common predictive measures (Author Material XI-D), it may represent a more unique construct in the literature.

This study had multiple limitations. Though one of the largest prognostic fMRI studies in psychiatry, our samples were small by the standards of clinical trials. Randomized trials are needed to make testable inferences regarding the relationship of sgACC function specifically to CT outcome. Future examination of medicated depressed participants will help to generalize to clinically representative conditions. If healthy individuals' moderate reliability ( $r < .5$ ) is representative of patients, refinement in multiple-baseline patient studies is needed before inferring clinical applicability to single patients. Amygdala results were not consistent with<sup>4</sup>, possibly reflecting our use of an anatomically defined amygdala region versus the empirically derived smaller region in<sup>4</sup>. Our index captured sustained processing of emotional information but could have missed preparatory or initial aspects of reactivity.

These limitations notwithstanding, our data suggest that a simple, automated, and theoretically motivated neuroimaging assessment yields a marker of whether an individual is likely to respond to a specific validated treatment for unipolar depression, Cognitive Therapy. Though the data do not suggest that neuroimaging should be used clinically at this time, they do indicate that neuroimaging is ready for "next step" investigations, including larger randomized clinical trials, creating age, gender, and scanner, norms, and considerations regarding dissemination.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We gratefully acknowledge the contributions of Agnes Haggerty, Mauri Cesare, and the staff of the Mood Disorders Treatment and Research Program at Western Psychiatric Institute and Clinic, and the intrepid community therapists who worked on the study. We thank our patients, without whom none of this work would have been possible.

**Supported by** This research was supported by the National Institutes of Health MH074807, MH082998, MH58356, MH58397, MH69618, and the Pittsburgh Foundation, Emmerling Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. ClinicalTrials.gov, identifiers: NCT00183664, NCT00787501

## References

1. Beck, AT.; Rush, AJ.; Shaw, BF.; Emery, G. Cognitive therapy of depression. New York: Guilford Press; 1979.
2. Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychological Science in the Public Interest*. 2002; 3(2):39–77.
3. Kemp AH, Gordon E, Rush AJ, Williams LM. Improving the prediction of treatment response in depression: integration of clinical, cognitive, psychophysiological, neuroimaging, and genetic measures. *CNS Spectr*. Dec; 2008 13(12):1066–1086. quiz 1087–1068. [PubMed: 19179943]
4. Siegler GJ, Carter CS, Thase ME. Use of fMRI to predict recovery from unipolar depression with Cognitive Behavior Therapy. *Am J Psychiatry*. Apr; 2006 163(4):735–738. [PubMed: 16585452]

5. Fu CH, Williams SC, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, Donaldson C, Suckling J, Andrew C, Steiner H, Murray RM. Neural Responses to Sad Facial Expressions in Major Depression Following Cognitive Behavioral Therapy. *Biol Psychiatry*. Jun 10.2008
6. Bruder GE, Stewart JW, Mercier MA, Agosti V, Leite P, Donovan S, Quitkin FM. Outcome of cognitive-behavioral therapy for depression: relation to hemispheric dominance for verbal processing. *J Abnorm Psychol*. Feb; 1997 106(1):138–144. [PubMed: 9103725]
7. Mayberg HS. Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimised treatment. *British Medical Bulletin*. 2003; 65:193–207. [PubMed: 12697626]
8. Mayberg HS, Liotti M, Brannan SK, McGinnis BS, Mahurin RK, Jerabek PA, Silva JA, Janet LT, Martin CC, Lancaster JL, Fox PT. Reciprocal limbic cortical function and negative mood Converging PET findings in depression and normal sadness. *American Journal of Psychiatry*. 1999; 156:675–682. [PubMed: 10327898]
9. Drevets WC. Neuroimaging studies of mood disorders. *Biological Psychiatry*. 2000; 48(8):813–829. [PubMed: 11063977]
10. Fu CH, Williams SC, Cleare AJ, Brammer MJ, Walsh ND, Kim J, Andrew CM, Pich EM, Williams PM, Reed LJ, Mitterschiffthaler MT, Suckling J, Bullmore ET. Attenuation of the neural response to sad faces in major depression by antidepressant treatment: a prospective, event-related functional magnetic resonance imaging study. *Arch Gen Psychiatry*. Sep; 2004 61(9):877–889. [PubMed: 15351766]
11. Goldapple K, Segal Z, Garson C, Lau M, Bieling P, Kennedy S, Mayberg H. Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Arch Gen Psychiatry*. Jan; 2004 61(1):34–41. [PubMed: 14706942]
12. Mayberg HS, Brannan SK, Mahurin RK, Jerabek PA, Brickman JS, Tekell JL. Cingulate function in depression: A potential predictor of treatment response. *Neuroreport*. 1997; 8(4):1057–1061. [PubMed: 9141092]
13. Keedwell PA, Drapier D, Surguladze S, Giampietro V, Brammer M, Phillips M. Subgenual cingulate and visual cortex responses to sad faces predict clinical outcome during antidepressant treatment for depression. *J Affect Disord*. Jun 17.2009
14. Bryant RA, Felmingham K, Kemp A, Das P, Hughes G, Peduto A, Williams L. Amygdala and ventral anterior cingulate activation predicts treatment response to cognitive behaviour therapy for post-traumatic stress disorder. *Psychol Med*. Apr; 2008 38(4):555–561. [PubMed: 18005496]
15. Vul E, Harris C, Winkelman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*. 2009; 4(3): 274–290.
16. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*. May; 2009 12(5):535–540. [PubMed: 19396166]
17. DeRubeis RJ, Siegle GJ, Hollon SD. Cognitive therapy versus medication for depression: treatment outcomes and neural mechanisms. *Nat Rev Neurosci*. Oct; 2008 9(10):788–796. [PubMed: 18784657]
18. Canli T, Cooney RE, Goldin P, Shah M, Sivers H, Thomason ME, Whitfield-Gabrieli S, Gabrieli JD, Gotlib IH. Amygdala reactivity to emotional faces predicts improvement in major depression. *Neuroreport*. Aug 22; 2005 16(12):1267–1270. [PubMed: 16056122]
19. Davidson RJ. Assymetric brain function affective style and psychopathology: The role of early experience and plasticity. *Development and Psychopathology*. 1994; 6:741–758.
20. Siegle GJ, Thompson W, Carter CS, Steinhauer SR, Thase ME. Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: Related and independent features. *Biol Psychiatry*. Oct 5; 2007 61(2):198–209. [PubMed: 17027931]
21. Pizzagalli DA, Peccoralo LA, Davidson RJ, Cohen JD. Resting anterior cingulate activity and abnormal responses to errors in subjects with elevated depressive symptoms: a 128-channel EEG study. *Hum Brain Mapp*. Mar; 2006 27(3):185–201. [PubMed: 16035102]

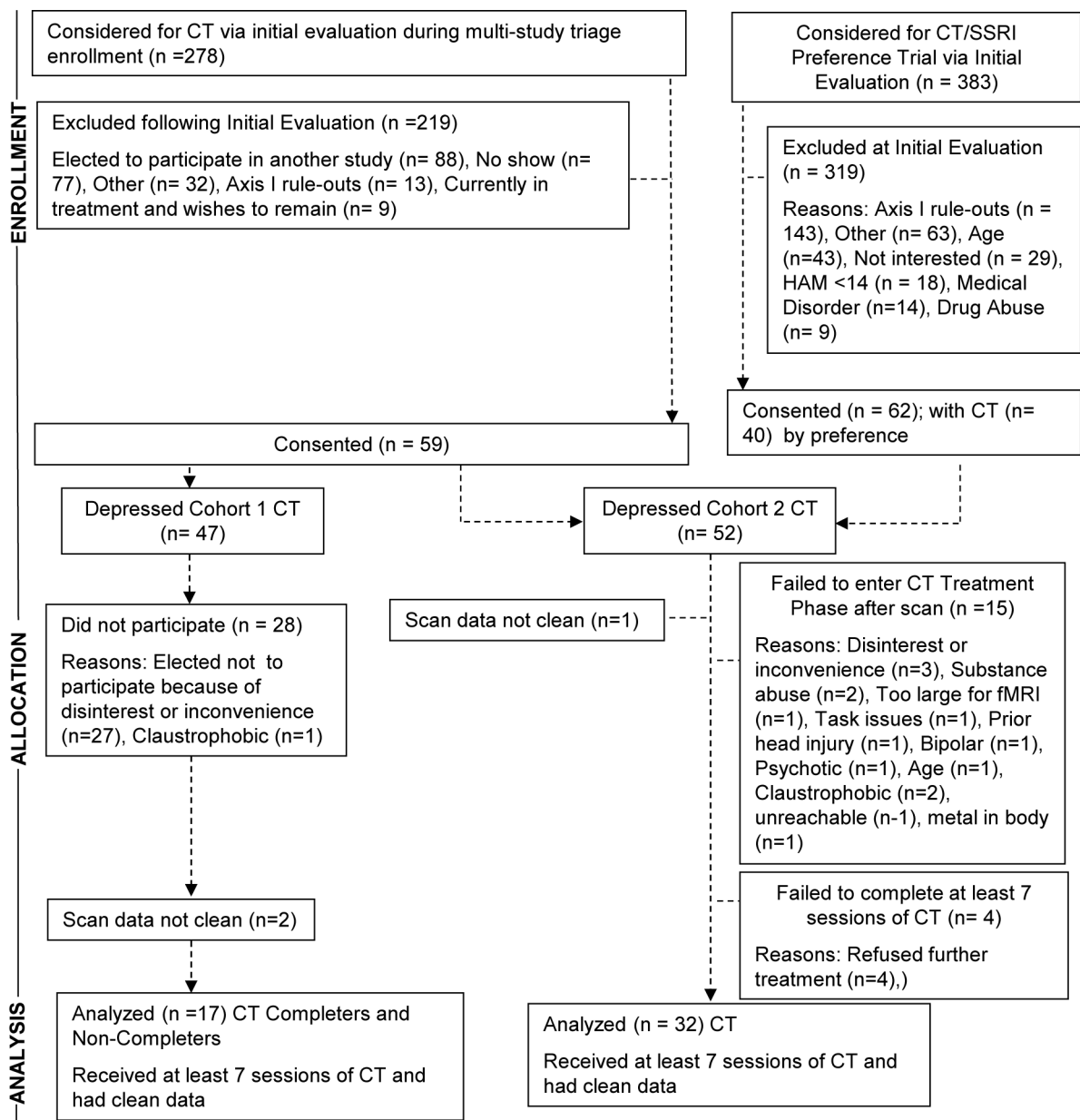
22. Lemogne C, le Bastard G, Mayberg H, Volle E, Bergouignan L, Lehericy S, Allilaire JF, Fossati P. In search of the depressive self: extended medial prefrontal network during self-referential processing in major depression. *Soc Cogn Affect Neurosci*. Mar 23.2009
23. Dichter GS, Felder JN, Smoski MJ. Affective context interferes with cognitive control in unipolar depression: an fMRI investigation. *J Affect Disord*. Apr; 2009 114(1–3):131–142. [PubMed: 18706701]
24. Davidson RJ. Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*. Sep; 2003 40(5):655–665. [PubMed: 14696720]
25. Anand A, Li Y, Wang Y, Wu J, Gao S, Bukhari L, Mathews VP, Kalnin A, Lowe MJ. Activity and connectivity of brain mood regulating circuit in depression: a functional magnetic resonance study. *Biol Psychiatry*. May 15; 2005 57(10):1079–1088. [PubMed: 15866546]
26. Cooney RE, Joormann J, Atlas LY, Eugene F, Gotlib IH. Remembering the good times: neural correlates of affect regulation. *Neuroreport*. Nov 19; 2007 18(17):1771–1774. [PubMed: 18090309]
27. Campbell-Sills L, Simmons AN, Lovero KL, Rochlin AA, Paulus MP, Stein MB. Functioning of neural systems supporting emotion regulation in anxiety-prone individuals. *NeuroImage*. Jan 1; 2011 54(1):689–696. [PubMed: 20673804]
28. Drevets WC, Savitz J, Trimble M. The subgenual anterior cingulate cortex in mood disorders. *CNS Spectr*. Aug; 2008 13(8):663–681. [PubMed: 18704022]
29. Hamilton JP, Glover GH, Hsu JJ, Johnson RF, Gotlib IH. Modulation of subgenual anterior cingulate cortex activity with real-time neurofeedback. *Hum Brain Mapp*. Jan; 2011 32(1):22–31. [PubMed: 21157877]
30. Liotti M, Mayberg HS, McGinnis S, Brannan SL, Jerabek P. Unmasking disease-specific cerebral blood flow abnormalities: mood challenge in patients with remitted unipolar depression. *American Journal of Psychiatry*. 2002; 159(11):1830–1840. [PubMed: 12411216]
31. Matthews S, Simmons A, Strigo I, Gianaros P, Yang T, Paulus M. Inhibition-related activity in subgenual cingulate is associated with symptom severity in major depression. *Psychiatry Res*. Apr 30; 2009 172(1):1–6. [PubMed: 19239982]
32. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C, Schwab JM, Kennedy SH. Deep brain stimulation for treatment-resistant depression. *Neuron*. Mar 3; 2005 45(5):651–660. [PubMed: 15748841]
33. Siegler GJ. Beyond depression commentary: Wherefore art thou, Depression Clinic of Tomorrow? *Clinical Psychology Science and Practice*. 2011; 18:305–310.
34. First, MB.; Spitzer, RL.; Gibbon, M.; Williams, JB. Structured Clinical Interview for DSM IV Axis I Disorders Patient Edition. Vol. 20. New York: Biometrics Research Department New York State Psychiatric Institute; 1996.
35. Young, J.; Beck, AT. Cognitive Therapy Scale: Rating Manual. Philadelphia: Center for Cognitive Therapy; 1980.
36. Nelson, HE.; Willison, J. The National Adult Reading Test. Berkshire, England: Nfer-Nelson Publishing Co; 1991.
37. Beck, AT.; Steer, RA.; Brown, G. Beck Depression Inventory Second Edition Manual. San Antonio: The Psychological Corporation; 1996.
38. Bradley, MM.; Lang, PJ. Affective Norms for English Words ANEW Technical Manual and Affective Ratings. Gainesville FL: The Center for Research in Psychophysiology University of Florida; 1997.
39. Siegler GJ, Granholm E, Ingram RE, Matt GE. Pupillary response and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry*. 2001; 49:624–636. [PubMed: 11297720]
40. Siegler GJ, Steinhauer SR, Carter CS, Ramel W, Thase ME. Do the seconds turn into hours? Relationships between sustained pupil dilation in response to emotional information and self reported rumination. *Cognitive Therapy and Research*. 2003; 27(3):365–383.
41. Siegler GJ, Steinhauer SR, Thase ME, Stenger VA, Carter CS. Can't shake that feeling: fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biological Psychiatry*. 2002; 51:693–707. [PubMed: 11983183]

42. Breiman L. Random forests. *Machine Learning*. 2001; 45:5–21.
43. Zhou D, Thompson WK, Siegle G. MATLAB toolbox for functional connectivity. *Neuroimage*. Oct 1; 2009 47(4):1590–1607. [PubMed: 19520177]
44. Fu CH, Williams SC, Cleare AJ, Scott J, Mitterschiffthaler MT, Walsh ND, Donaldson C, Suckling J, Andrew C, Steiner H, Murray RM. Neural responses to sad facial expressions in major depression following cognitive behavioral therapy. *Biol Psychiatry*. Sep 15; 2008 64(6):505–512. [PubMed: 18550030]
45. Phillips ML, Ladouceur CD, Drevets WC. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol Psychiatry*. Sep; 2008 13(9):829, 833–857. [PubMed: 18574483]
46. Brannan SK, Mayberg HS, McGinnis S, Silva JA, Tekell J, Mahurin RK, Jerabek PA, Fox P. Cingulate metabolism predicts treatment response: A replication. *Biological Psychiatry*. 2000; 47:107S. #355.
47. Brody AL, Saxena S, Mandelkern MA, Fairbanks LA, Ho ML, Baxter LR. Brain metabolic changes associated with symptom factor improvement in major depressive disorder. *Biol Psychiatry*. Aug 1; 2001 50(3):171–178. [PubMed: 11513815]
48. Davidson RJ, Irwin W, Anderle MJ, Kalin NH. The neural substrates of affective processing in depressed patients treated with venlafaxine. *American Journal of Psychiatry*. 2003; 160(1):64–75. [PubMed: 12505803]
49. Chen CH, Ridler K, Suckling J, Williams S, Fu CH, Merlo-Pich E, Bullmore E. Brain Imaging Correlates of Depressive Symptom Severity and Predictors of Symptom Improvement After Antidepressant Treatment. *Biol Psychiatry*. Jan 8.2007
50. Nitschke JB, Sarinopoulos I, Oathes DJ, Johnstone T, Whalen PJ, Davidson RJ, Kalin NH. Anticipatory activation in the amygdala and anterior cingulate in generalized anxiety disorder and prediction of treatment response. *Am J Psychiatry*. Mar; 2009 166(3):302–310. [PubMed: 19122007]
51. Langenecker SA, Kennedy SE, Guidotti LM, Briceno EM, Own LS, Hooven T, Young EA, Akil H, Noll DC, Zubieta JK. Frontal and limbic activation during inhibitory control predicts treatment response in major depressive disorder. *Biol Psychiatry*. Dec 1; 2007 62(11):1272–1280. [PubMed: 17585888]
52. Roy M, Harvey PO, Berlim MT, Mamdani F, Beaulieu MM, Turecki G, Lepage M. Medial prefrontal cortex activity during memory encoding of pictures and its relation to symptomatic improvement after citalopram treatment in patients with major depression. *J Psychiatry Neurosci*. May; 2010 35(3):152–162. [PubMed: 20420766]
53. Pizzagalli D, Pascual-Marqui RD, Nitschke JB, Oakes TR, Larson CL, Abercrombie HC, Schaefer SM, Koger JV, Benca RM, Davidson RJ. Anterior cingulate activity as a predictor of degree of treatment response in major depression: evidence from brain electrical tomography analysis. *Am J Psychiatry*. Mar; 2001 158(3):405–415. [PubMed: 11229981]
54. Siegle GJ, Steinhauer SR, Friedman ES, Thompson WS, Thase ME. Remission prognosis for cognitive therapy for recurrent depression using the pupil: utility and neural correlates. *Biol Psychiatry*. Apr 15; 2011 69(8):726–733. [PubMed: 21447417]

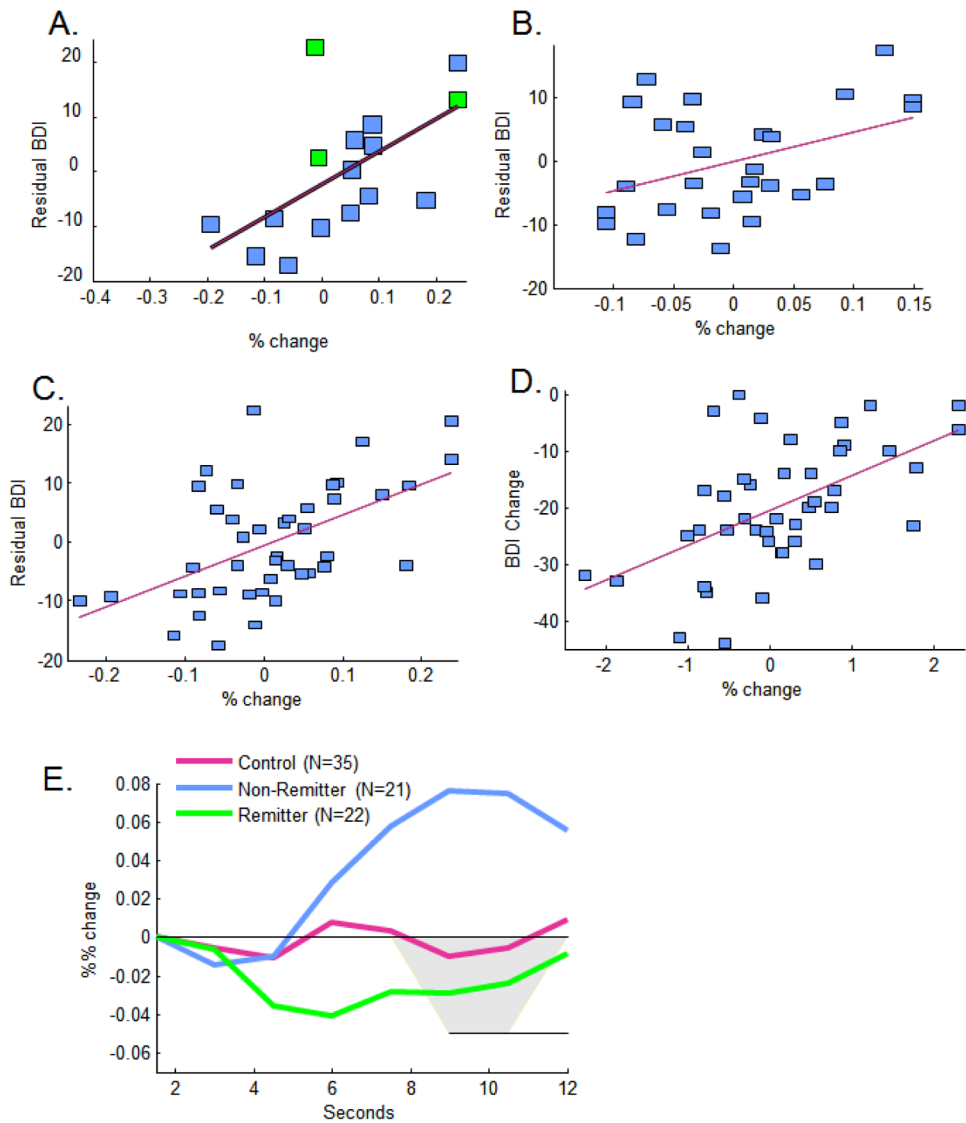


ClinicalTrials.gov: NCT00183664

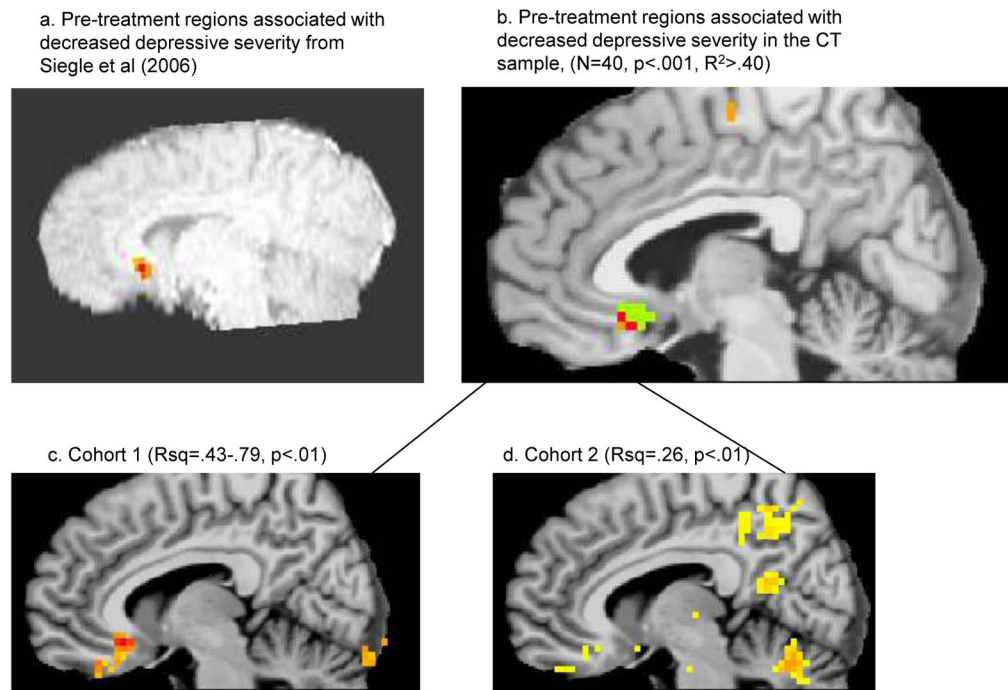
ClinicalTrials.Gov: NCT00787501



**Figure 1.**  
CONSORT diagram

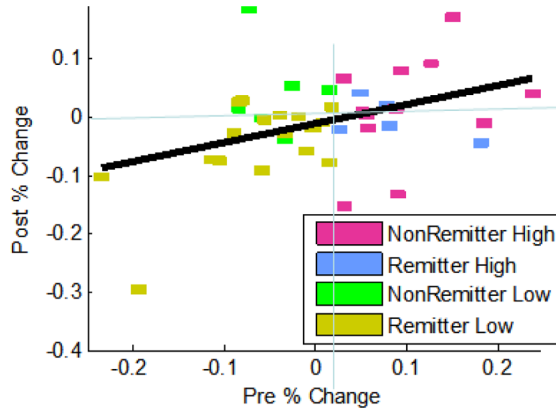


**Figure 2.** Decreased anatomically defined sgACC activity was strongly correlated with stronger clinical response (decreased residual BDI) in A) Cohort 1 (green are non-completers whose final BDI scores were interpolated), moderately associated in B) Cohort 2, and more strongly associated in C) the combined cohort as well as D) the combined cohort using Z scores for sgACC and change in BDI as the outcome variable. E) Waveforms for hemodynamic responses for nonremitters (final BDI  $\geq 10$ ; N=22) compared to controls (N=35) and remitters (final BDI  $< 10$ ; N=22). Areas significant at  $p < .05$  upon ANOVAs at each scan are highlighted in grey.

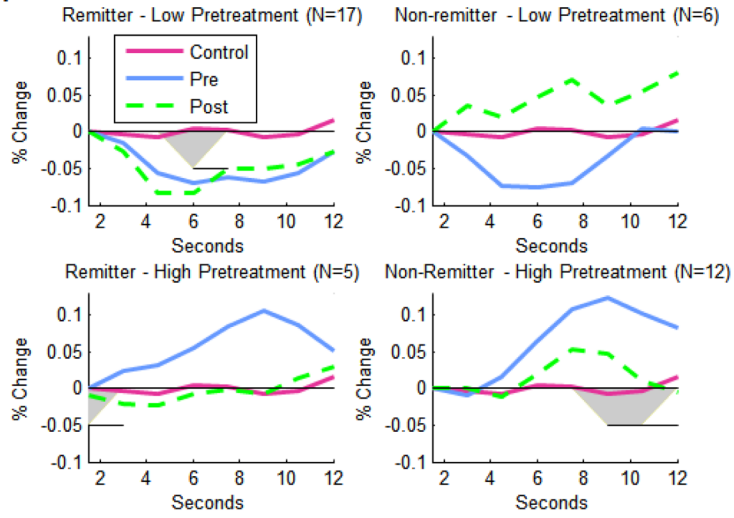


**Figure 3.** Decreased empirically defined sgACC activity associated strongly with response in A) our previous study<sup>4</sup> and B) the new samples ( $R^2 > .4$ ) – green is the anatomical ROI used as a mask, orange regions were only in the new dataset, and red voxels are overlapping. These regions were reflected in predictive regions in both (C) Cohort 1 and (D) Cohort 2.

A.



B.



**Figure 4.** Relationships of pre-treatment sgACC reactivity to post-treatment sgACC activity, A) continuously for the temporal region of interest, and B) with regard to waveforms in participants who were predicted to remit (pre-treatment %-change <.02) or not remit. Areas significant at  $p < .05$  upon ANOVAs at each scan are highlighted in grey.

Table 1

Demographic and clinical characteristics. T-tests confirmed that no groups or cohorts were different from any other on any demographic variable below (gender, ethnicity, age range, age, education). Cohorts did not differ in depressive severity or median # of depressive episodes. Group and cohort differences in emotion ratings and reaction times are described in the text and Author Material-VI. For BDI and HRSD, Response was defined as a 50% decrease in BDI or Hamilton. The first lines, of the clinical response section, using only CT completers, replicates the analysis done in Siegler et al (2006). Remaining lines generalize that result. Remission was defined as BDI<10 or HRSD <10

Measure	Depressed		Depressed Cohort 2 CT	Control		Test
	Cohort 1 CT	Cohort 2 CT		Cohort 1	Cohort 2	
N	17	32	15	20		
# male	4	5	7	5		
# Caucasian	14	29	12	17		
Age range	24-51	19-55	25-54	21-52		
Age M(SD)	37.41 (9.17)	36.13 (11.11)	38.93 (9.74)	32.95 (9.37)		
Years Education M(SD)	15.35 (2.23)	15.06 (2.60)	15.67 (2.26)	16.30 (2.13)		
Median # Depressive Episodes	5.59 (3.52)	4.03 (4.83)	0	0		
NAART VIQ equivalent	110.31 (8.68)	108.42 (9.92)	111.61 (7.97)	109.21 (8.96)		
Emotion ratings for Negative words (1=very negative-7=very positive) M(SD)	2.55 (0.66)	2.11 (0.62)	2.99 (0.65)	2.49 (0.72)		t(82)=-2.89, p=0.005, d=0.64
RT to negative words M(SD) in ms	1403.82 (269.41)	1637.34 (526.16)	1411.31 (390.26)	1242.20 (374.90)		t(82)=-2.52, p=0.01, d=0.56
Personal relevance ratings for Negative Words (1=not relevant, 5=very relevant) M(SD)	3.07(.66)	3.03(.57)	2.33(.72)	2.07(.49)		t(82)=6.57, p<.0005, d=1.43
Clinical Change Completers only N=13 as in (11)						
Pre-treatment BDI M(SD)	33.0 (8.8)					
Post-treatment BDI M(SD) Completers only	10.7 (10.2)					
BDI Change M(SD)	22.3 (13.4)					
N Responders	9 (69%)					
N Remitters	8 (61%)					
Clinical Change – whole sample: N BDI	16	27	15	20		
Pre-treatment BDI M(SD)	34.4 (8.7)	30.7 (9.2)	1.87 (3.58)	1.26 (1.63)		
Post-Treatment BDI M(SD)	14.1 (12.2)	11.3 (9.6)				
Change BDI M(SD)	20.3 (13.4)	19.4 (9.7)				
N Responders BDI	10 (63%)	17 (62%)				



Measure	Depressed	Control	Control	Control	Test
	Cohort 1 CT	Depressed Cohort 2 CT	Cohort 1	Cohort 2	Depressed v. Control
N Remitters BDI	8 (50%)	14 (51%)			
Clinical change – whole sample N (HRSD)	17	32			
Pre-treatment HRSD M(SD)	24.6 (5.7)	20.6 (4.8)			
Post-Treatment HRSD M(SD)	9.6 (6.5)	8.1 (5.8)			
Change HRSD M(SD)	15.0 (7.7)	12.5 (5.6)			
N Responders HRSD M(SD)	10 (63%)	20 (63%)			
N Remitters HRSD M(SD)	7 (47%)	13 (41%)			

**Table 2**

Associations of A) anatomical sgACC activity with clinical change using each cohort as a sample of interest and B) anatomical sgACC and a broader *a priori* network activity with clinical change, along with C) prediction for Cohort 2 using Cohort 1 as a training set. Note: a more elaborated version of this table including full Receiver Operating Characteristic (ROC) analyses is included in Author Material VIII-A, Table S2A.

Cohort	Measure	Residual	Response Prediction (% correct, permutation p)	Remission prediction (% correct, permutation p)
A. Prediction using sgACC with each Cohort as a sample of interest				
Cohort 1 <u>Completers</u> (as in <sup>4</sup> ) (N=14)	BDI	$R^2 = 0.414, F(1,12)=7.781, p=0.018$	84% correct, $p=.1$ (.04 based on $d'$ ), Sensitivity = 100%, Specificity = 86%, $d'=2.2$	84% correct, $p=.08$ (.04 based on $d'$ ), Sensitivity = 100%, Specificity = 75%, $d'=2.1$
Cohort 1 <u>Completers and Noncompleters</u> (estimated) (N=16)	BDI	$R^2 = 0.358, F(1,15)=7.79, p=0.014$	81% correct, $p=.07$ , Sensitivity = 83%, Specificity = 80%, $d'=1.8$	75% correct, $p=.14$ , Sensitivity = 100%, Specificity = 50%, $d'=1.4$
(N=17)	HRSD	$R^2 = 0.341, F(1,16)=7.772, p=0.014$	76% correct, $p=.41$ , Sensitivity = 80%, Specificity = 67%, $d'=1.3$	82% correct, $p=.045$ , Sensitivity = 100%, Specificity = 57%, $d'=1.7$
Cohort 2 (N=27)	BDI	$R^2 = 0.167, F(1,26)=4.91, p=0.04$	78% correct, $p=.04$ , Sensitivity = 40%, Specificity = 100%, $d'=1.5$	74% correct, $p=.04$ , Sensitivity = 54%, Specificity = 93%, $d'=1.6$
(N=32)	HRSD	$R^2 = 0.040, F(1,31)=1.244, p=0.273$	78% correct, $p=.02$ , Sensitivity = 42%, Specificity = 100%, $d'=1.61$	66% correct, $p=.38$ , Sensitivity = 32%, Specificity = 100%, $d'=1.32$
Cohorts 1 & 2 combined (N=43 with pre/post BDI scores (of N=49)	BDI	$R^2 = 0.288, F(1,42)=16.61, p<0.0005$	79% correct, $p=.003$ , Sensitivity = 50%, Specificity = 96%, $d'=1.79$	72% correct, $p=.01$ , Sensitivity = 38%, Specificity = 95%, $d'=1.39$
(N=49)	HRSD	$R^2 = 0.12, F(1,48)=6.29, p=0.02$	78% correct, $p=.001$ , Sensitivity = 47%, Specificity = 93.8%, $d'=1.46$	69% correct, $p=.06$ , Sensitivity = 34%, Specificity = 100%, $d'=1.58$
B. Prediction of residual symptoms using a priori network with combined cohort as a sample of interest (N=43) *				
Zero order relationships All $p$ 's <.05 – see Author Material VIII Table S2a for statistics.	BDI	sg ACC	$R^2 = 0.29$	
		R Amygdala	$R^2 = 0.16$	
		L DLPFC	$R^2 = 0.20$	
		BA24 in the VM PFC	$R^2 = 0.11$	
Multivariate relationships Full model $R^2=.43, F(1,38)=7.39, p=.0001$ . Statistics in Author Material VIII Table S2a	BDI	Constant	$s\beta=0, p=.21$	
		sgACC	$s\beta=.47, p=.0005$	
		R Amygdala	$s\beta=.24, p=.10$	
		L DLPFC	$s\beta=.30, p=.05$	

Cohort	Measure	Residual	Response Prediction (% correct, permutation p)		Remission prediction (% correct, permutation p)	
C. Generalizable classification from just sgACC or all 4 regions and their connectivity using random forest methodology with Cohort 2 as the test set.***						
1, CT – training set – all brain regions	BDI	R <sup>2</sup> = 0.28, p<0.02	75% correct, Sensitivity = 75%, Specificity = 90%, d' = 1.28, p = .023	69% correct, Sensitivity = 62%, Specificity = 75%, d' = .99, p = .057		
	HRSD	R <sup>2</sup> = 0.21, p<0.01	76% correct, Sensitivity = 92%, Specificity = 40%, d' = 1.13, p = .028	65% correct, Sensitivity = 57%, Specificity = 70%, d' = .7, p = .12		
2, CT – Generalization set – initial severity alone	BDI	R <sup>2</sup> = 0, p<0.74	48% correct, Sensitivity = 53%, Specificity = 40%, d' = -.18, p = .63	38% correct, Sensitivity = 23%, Specificity = 54%, d' = -.64, p = .92		
	HRSD	R <sup>2</sup> = 0, p = 0.63	56% correct, Sensitivity = 60%, Specificity = 50%, d' = .25, p = .29	47% correct, Sensitivity = 38%, Specificity = 53%, d' = -.23, p = .7		
2, CT - Generalization set – initial severity + sgACC	BDI	R <sup>2</sup> = 0, p = 0.07	74% correct, Sensitivity = 100%, Specificity = 30%, d' = 1.04, p = .03	78% correct, Sensitivity = 86%, Specificity = 69%, d' = 1.57, p = .003		
	HRSD	R <sup>2</sup> = 0.02, p = .07	81% correct, Sensitivity = 95%, Specificity = 58%, d' = 1.86, p < .001	62% correct, Sensitivity = 69%, Specificity = 58%, d' = .7, p = .08		
2, CT – Generalization set – initial severity + all brain regions (retained: Baseline BDI, LDLPFC, sgACC, phi_LDLPFC_sgACC)	BDI	R <sup>2</sup> = 0.15, p < 0.003	81% correct, Sensitivity = 100%, Specificity = 44%, d' = 1.43, p = .0005	63% correct, Sensitivity = 79%, Specificity = 46%, d' = .7, p = .086		
	HRSD	R <sup>2</sup> = 0.07, p = .02, M(points off) = 3.8	75% correct, Sensitivity = 95%, Specificity = 42%, d' = 1.43, p = .002	66% correct, Sensitivity = 85%, Specificity = 53%, d' = 1.09, p = .015		

\* Classification not evaluated in the multivariate model without robust estimation (2B) given the potential for type I error – rather, evaluations in 2C reflect robust estimations from the multivariate model.

\*\*\* P-values for random forests were estimated via permutation tests on %-correct using the same random-forest methodology. Regression and classification accuracy estimates unbiased computed on out-of-bag samples for robust generalization.